UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/844,040	04/27/2001	Brian Seong-Gon Kim	089984-0276267	2687

27498   7590   12/22/2003

PILLSBURY WINTHROP LLP
2475 HANOVER STREET
PALO ALTO, CA 94304-1114

EXAMINER
CHEN, CHONGSHAN

ART UNIT	PAPER NUMBER
2172	

DATE MAILED: 12/22/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 10/03)

-- *The MAILING DATE of this communication appears on the cover sheet with the correspondence address* --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on <u>26 September 2003</u>.

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) <u>1-39</u> is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) <u>1-39</u> is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. §§ 119 and 120

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

13)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application) since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.

    a) ☐ The translation of the foreign language provisional application has been received.

14)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121 since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.

Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☐ Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____.

4)☐ Interview Summary (PTO-413) Paper No(s). _____.

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: .

DETAILED ACTION

1.   This action is responsive to communications: Amendment A, filed on 26 September

2003. Claims 1-39 are pending.

Claim Rejections - 35 USC § 112

2.   The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

3.   Claims 2, 17, 20 and 26 rejected under 35 U.S.C. 112, second paragraph, as being

indefinite for failing to particularly point out and distinctly claim the subject matter which

applicant regards as the invention.

The phrase "determine if there are enough documents", "enough documents", "where

there is too much overlap" and so on are indefinite. How many is enough? How much is

too much?

Claim Rejections - 35 USC § 103

4.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

5.   Claims 1-16 and 18-39 are rejected under 35 U.S.C. 103(a) as being unpatentable over

Ferguson et al. ("Ferguson", 6,237,011) in view of Ho et al. ("Ho", 6,571,240).

As per claim 1, Ferguson teaches a method of categorizing an initial collection of documents, each document being represented by a string of characters, the method comprising the steps of:

identifying predefined characters in the string of characters from the documents in the initial collection of documents to form identified characters (Ferguson, col. 8, lines 12-21, "… the key words and/or attributes are automatically extracted …");

constructing a number of categories from the string of characters of the preprocessed collection of documents (Ferguson, col. 8, lines 12-32); and

assigning each document in the preprocessed collection of documents to a category to form a hierarchy of categories of documents (Ferguson, col. 8, lines 12-32, "categorizing a document into one or more categories …").

Ferguson does not explicitly disclose changing the identified characters in the documents in the initial collection of documents to form a preprocessed collection of documents, each of the preprocessed collection of documents represented by a preprocessed string of characters. Ho teaches changing the identified characters in the documents (Ho, col. 12, lines 16-26). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to change the identified characters in the documents in the system of Ferguson. Because the key words in the document might be in a non-regularizable form, e.g. non-root form, this provides difficulty for the system to match key words and find similarity between documents. The regularizer of Ho will regularize the key words in a document in order to allow the system easily to match key words.

As per claim 2, Ferguson and Ho teach all the claimed subject matters as discussed in claim 1, and further teach

clearing a temporary category and selecting a seed document from the preprocessed collection of documents as a first document of the temporary category (Ferguson, col. 8, lines 12-32);

collecting documents from the preprocessed collection of documents that are similar to the seed document into the temporary category (Ferguson, col. 8, lines 12-32);

testing to determine if there are enough documents in the temporary category to merit construction of a new category (Ferguson, col. 8, lines 12-32);

constructing the new category from the temporary category and generating a heading for the new category if there are enough documents in the temporary category to merit construction and generation (Ferguson, col. 8, lines 12-32);

assigning the seed document to a category reserved for documents not belonging to any specific category if there are not enough documents in the temporary category (Ferguson, col. 8, lines 12-32); and

marking the documents assigned to any category in the preprocessed collection of documents as processed (Ferguson, col. 8, lines 12-32).

As per claim 3, Ferguson and Ho teach all the claimed subject matters as discussed in claim 2, except for explicitly disclosing the predefined characters include punctuation marks, and the changing step removes the punctuation marks from the string of characters. However, Ho discloses a document regularizer regularizes words or phrase in the document (Ho, col. 12, lines 16-26). It would have been obvious to one of ordinary skill in the art at the time the invention

was made to use the regularizer to remove the punctuation marks from the string of characters because the punctuation marks will give difficulty to the categorization system to match key words.

Claim 4 is rejected on grounds corresponding to the reasons given above for claim 3.

As per claim 5, Ferguson and Ho teach all the claimed subject matters as discussed in claim 2, and further teach predefined characters include non-root words, and the changing step replaces the non-root words with root words (Ho, col. 12, lines 16-26).

Claims 6-7 are rejected on grounds corresponding to the reasons given above for claim 3.

As per claim 8, Ferguson and Ho teach all the claimed subject matters as discussed in claim 2, and further teach loading a character string from the seed document into a memory location to initialize the values of a number of category properties for the temporary category (Ferguson, col. 8, lines 12-32).

As per claim 9, Ferguson and Ho teach all the claimed subject matters as discussed in claim 8, and further teach

determining if there are documents in the preprocessed collection of documents that have not been processed with respect to the temporary category (Ferguson, col. 8, lines 12-32);

if there are documents in the preprocessed collection of documents that have not been processed with respect to the temporary category, selecting a next document from the preprocessed collection of documents and measuring a similarity of the preprocessed string of characters of the text document using a similarity test between the next document and the values of the number of current category properties (Ferguson, col. 8, lines 12-32);

including the next document in the temporary category if the next document passes the

similarity test (Ferguson, col. 8, lines 12-32);

updating the values of the number of category properties of the temporary category when

the next document is included (Ferguson, col. 8, lines 12-32); and

rejecting the next document if the next document fails the similarity test (Ferguson, col.

8, lines 12-32).

As per claim 10, Ferguson and Ho teach all the claimed subject matters as discussed in

claim 9, and further teach repeating the steps of claim 9 for all documents in preprocessed

collection of documents (Ferguson, col. 8, lines 12-32).

As per claim 11, Ferguson and Ho teach all the claimed subject matters as discussed in

claim 2, and further teach collecting more similar documents from a number of existing

categories (Ferguson, col. 8, lines 12-32).

As per claim 12, Ferguson and Ho teach all the claimed subject matters as discussed in

claim 11, and further teach

determining if there are more documents in a number of existing categories that have not

been processed with respect to the temporary category (Ferguson, col. 8, lines 12-32);

if there are documents in the number of existing categories that have not been processed

with respect to the temporary category, selecting a next document from the number of existing

categories as a selected document and measuring a similarity of the preprocessed string of

characters of the selected document using a similarity test between the selected document and

values of a number of current category properties (Ferguson, col. 8, lines 12-32);

including the selected document in the temporary category if the selected document passes the similarity test (Ferguson, col. 8, lines 12-32); and

rejecting the selected document if the selected document fails the similarity test (Ferguson, col. 8, lines 12-32).

As per claim 13, Ferguson and Ho teach all the claimed subject matters as discussed in claim 12, and further teach repeating the steps of claim 12 for all documents in the number of existing categories (Ferguson, col. 8, lines 12-32).

As per claim 14, Ferguson and Ho teach all the claimed subject matters as discussed in claim 8, and further teach the category properties includes a string of characters selected from the group consisting of a longest common substring in the title, a longest common substring in the body; and a document type index measured as list of fractional numbers for each document type (Ferguson, col. 8, lines 12-32).

As per claim 15, Ferguson and Ho teach all the claimed subject matters as discussed in claim 14, and further teach categorizing documents into categories (Ferguson, col. 8, lines 12-32), the documents inherently includes news article, technical documents, and poems.

As per claim 16, Ferguson and Ho teach all the claimed subject matters as discussed in claim 2, and further teach making sub-categories if there are too many documents in a given category; and post-processing the number of categorized lists of documents (Ho, col. 9, line 62 – col. 10, line 9).

As per claim 18, Ferguson and Ho teach all the claimed subject matters as discussed in claim 2, and further teach the seed document is a first document in the preprocessed collection of documents (Ferguson, col. 8, lines 12-32).

As per claim 19, Ferguson and Ho teach all the claimed subject matters as discussed in claim 2, and further teach the seed document is a document with a highest rank value among the documents not marked as processed in the preprocessed collection of documents (Ferguson, col. 8, lines 12-32).

As per claim 20, Ferguson and Ho teach all the claimed subject matters as discussed in claim 2, and further teach the temporary category is tested to determine if there are enough documents in the temporary category to merit construction of a new category by accumulating the weight of each document when each document can contribute uniform weight or different weight based on the rank value of each document with higher ranked document given more weight (Ferguson, col. 8, lines 12-32).

As per claim 21, Ferguson and Ho teach all the claimed subject matters as discussed in claim 2, and further teach except for explicitly disclosing the heading is a longest common substring in a title. It would have been obvious to one of ordinary skill in the art at the time the invention was made to use the longest common substring in a title as category heading because the longest common phrase in the title best describes the topic of the category. Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to use the longest common substring in a title as category heading because the longest common phrase in the title best describes the topic of the category.

Claim 22 is rejected on grounds corresponding to the reasons given above for claim 21.

As per claim 23, Ferguson and Ho teach all the claimed subject matters as discussed in claim 1, and further teach determining if an anchor-text character string is available for the documents in the initial collection ~of documents; and attaching an anchor-text character string

to the string of characters that represents the documents in the initial collection of documents when the anchor text character string is available (Ferguson, col. 8, lines 12-32).

As per claim 24, Ferguson and Ho teach all the claimed subject matters as discussed in claim 23, and further teach the anchor-text character string is a text used most frequently by hypertext documents (Ferguson, col. 8, lines 12-32).

As per claim 25, Ferguson and Ho teach all the claimed subject matters as discussed in claim 23, and further teach the anchor-text character string is a text with a highest partial extrinsic rank value (Ferguson, col. 8, lines 12-32).

Claims 26-39 are rejected on grounds corresponding to the reasons given above for claims 1-16 and 18-25.

6.   Claim 17 is rejected under 35 U.S.C. 103(a) as being unpatentable over Ferguson et al. ("Ferguson", 6,237,011) in view of Ho et al. ("Ho", 6,571,240) and further in view of Zamir (Oren Zamir and Oren Etzioni, "Grouper: A Dynamic Clustering Interface to Web Search Results").

As per claim 16, Ferguson and Ho teach all the claimed subject matters as discussed in claim 2, except for explicitly disclosing merging two categories that each have a heading where there is too much overlap in the headings of the two categories; and promoting sub-categories to an upper level in a hierarchy when there are not enough categories in the upper level. Zamir teaches merging clusters (Zamir, page 4, line 1). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to merge clusters in the system of Ferguson in order to avoid overlap.

Response to Arguments

7.   Applicant's arguments with respect to claims 1-39 have been considered but are moot in

view of the new ground(s) of rejection.

Conclusion

Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Chongshan Chen whose telephone number is 703-305-8319.

The examiner can normally be reached on Monday - Friday (8:00 am - 4:30 pm).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, John E Breene can be reached on (703)305-9790. The fax phone number for the

organization where this application or proceeding is assigned is (703) 872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding

should be directed to the receptionist whose telephone number is (703)305-3900.

December 13, 2003

SHAHID ALAM
PRIMARY EXAMINER